Standards for the Electronic Transfer of Clinical Data: Progress, Promises, and the Conductor's Wand

Clement J. McDonald, M.D.

Indiana University School of Medicine
and the Regenstrief Institute for Health Care
Indianapolis, IN 46202

Introduction

Standards are the lubricants of technical and economic progress. Nuts and bolts with the same width and gauge easily thread together, telephones can be interchanged from home to home and city to city, and U.S. electric razors, mix-masters, and televisions can be plugged into any U.S. wall socket, all because of standards. In fact, some standards are so entwined in the fabric of our life they seem to emanate from nature. Standards can be applied to almost any aspect of a product, to its composition, its environmental tolerance, its life-span. But the ones of most interest to this discussion are interface standards, those that connect the components of an assemblage. The introductory standards are all examples.

Interface standards have important benefits. They make it possible to assemble complex systems from simple standard parts. Some standards, such as the tread standards for nuts and bolts, only provide cheaper or better products. Others, by enabling the connection of previously un-connectable components inspire entirely new inventions. The Music Instrument Digital Interface (MIDI) is an example.[1] It was developed to permit musicians to mix and match keyboards and music synthesizers from different manufacturers. But it has spawned computer programs that convert musical keyboard input into sheet music and synthetic orchestras that consist of many synthetic instruments controlled by one or more keyboards. By clearly defining the behavior of components, interface standards encapsulate that component so that the development of each component can then proceed independently of the whole. Standards permit diversity at the components while promoting uniformity at the interface.

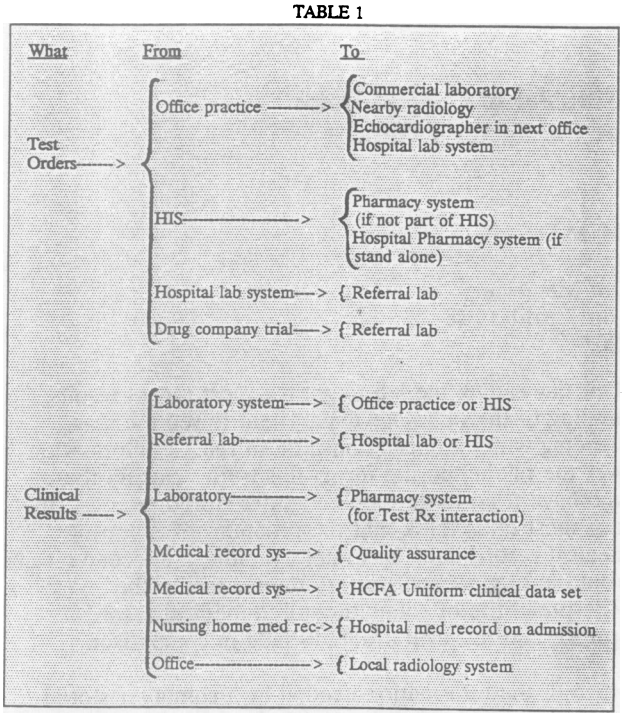
We first became interested in interface standards through our work with computer stored medical records. The medical record is an assemblage of information from "outside" sources: the clinical lab, the radiology department, the consultant, the nurse, the current physician as well as sources other than the current point of care. E.g., the discharge summary from an outside hospital, the records from a referring nursing home. Because people use the paper medical record most often after it has acquired patient information, they tend to forget the effort required to capture and assemble all of that data in the first place. Interns who spend much of their day calling the lab to get the most recent results, running to the radiology department to see the xray, pleading with an outside hospital to release information over the phone, do not forget this truth. Nor do those who have tried to develop computer stored records. In fact, most of us think that the problems of capturing medical record data, accurately, rapidly, and economically, from the original sources, are the *barriers* to the widespread use of computer stored medical records. Storage and retrieval problems pale in comparison.

Much of the information needed by a computer stored medical record is already stored electronically in ancillary service computers. Laboratory data resides in computerized lab systems, prescription information in pharmacy systems, and hospital discharge reports, surgical procedure report, consultant's notes in word processor computers, case abstract and order information on the hospital information system (HIS), and so on. Yet, for practical purposes this information is not available to a medical record or any other computer system. This is so because the representation of the data, the storage structures, the record identifiers, and the meanings of codes vary so much from vendor to vendor, and even from installation to installation within one vendor's product. Like the ocean castaway we see "water, water everywhere, and not a drop to drink." Hence the need for clinical data interchange standards (CDI). With CDI standards, medical record computers could tap into source computers and obtain their clinical contents in an "understandable" way. Such standards will also enable important data transfers among other kinds of clinical systems. (See examples in Table 1)

Ultimately such standards could increase the efficiency of health care institutions greatly. Data entered at one system would be available for automatic re-use by other clinical systems and could thus eliminate the many layers of chart review: -- for quality assurance, infection control, dictation. Important test results could be appended automatically to the discharge summary, there should be no need to dictate them. The benefits apply whether an institution uses a single vendor or a multi-vendor computer system. Single vendor systems need standards to capture office practice data to the institution when a patient is admitted and visa versa when the patient is discharged. Clinical data interchange standards will also make it easy to obtain and pool data across institutions for clinical and policy research.

0195-4210/90/0000/0009$01.00 © 1990 SCAMC, Inc.

TABLE 1



Beginnings

My first attempts to stir interest in developing standards for CDI was in the form of an editorial. It said, in short, that the computer stored records would be impossible for most office based physicians, unless the medical industry developed standards for CDI. In the editorial, we called attention to the UPC code (the bar code on all grocery products) and applauded the "grocers'" foresight for developing the UPC standard. We argued that the medical profession should show similar foresight and develop CDI standards. The editorial met the first reviewer's sword in 1981. Three years and nine submissions later (including two re-tries following editorial maybes), it had not seen the printing press. The reviewers argued with us on many points, including those we had not made: e.g., that "clinical data is nothing like grocery stock", that "it was the supermarkets, not the 'grocers', who developed the UPC", and that "standards would be of no use because physicians did not have computers in their office and never would".

In 1983 we began another tack. A group of clinicians, laboratorians and computer scientists gathered as a task force of AAMSI (one of the forebearers of AMIA, whose meeting you are now attending) to formulate a draft standard. In order to speed closure, we limited the scope of the initial effort to the interchange of clinical laboratory results. We started with the clinical laboratory on the basis of a variant of Sutton's law - "that's where the data is". In 1984 the above mentioned editorial finally found an accepting journal - M.D. Computing, where, coincidentally, I was the editor.[2] In the fall of that year we presented our draft standard to an open SCAMC meeting. The participants' responses were a cacophony of disagreement and encouragement. Some argued that even the limited scope of the laboratory was an impossible task, others, that unless the standard covered all medical communications, it was unworthy. Still others said we should plod on. We have had an open discussion as part of SCAMC in every subsequent year, including this one.

In 1984 we also became subcommittee E31.11 of the American Society of Testing and Materials (ASTM). This was an important step because ASTM is one of the few qualified consensus standards forming groups and gave us the tools to develop a formal consensus, with proper procedures and policies. ASTM also indemnifies participants against constraint of trade suits. The standard was accepted by ASTM and published as 1238-88 in 1988.[3] Subsequently, we have expanded the scope to the transmission of all clinical data, and the standard has been put into fairly wide use. ASTM 1238 is currently being used or implemented by sixteen of the larger referral labs, including some of the largest, SmithKline Beecham, MetPath, and Mayo Clinic laboratories. It is being used at a number of university medical centers, including Columbia University, UCLA, Indiana University, and Duke University. Drug companies have also required the transmission of clinical trial data in ASTM format. HL7, a sister standard to be discussed later, is being implemented at over forty institutions.

Implementation

What is the standard? Its purpose is to enable the interchange of patient oriented data among a wide variety of computers systems. It deals with the application level, ISO level 7. Standards for the lower levels are being developed by industry at large. For now, any of a variety of lower level protocols, such as Kermit, TCP/IP, X.400, will do.

We intended this to be a "poor man's" transmission standard -- one that will run on current software, hardware, and communication media (including RS232 lines) within any operating system or language. Hence the interchange message is transmitted as a restricted ASCII characters set, and the "lines" of the message cannot be longer than 220. The message may be of unlimited length, however.

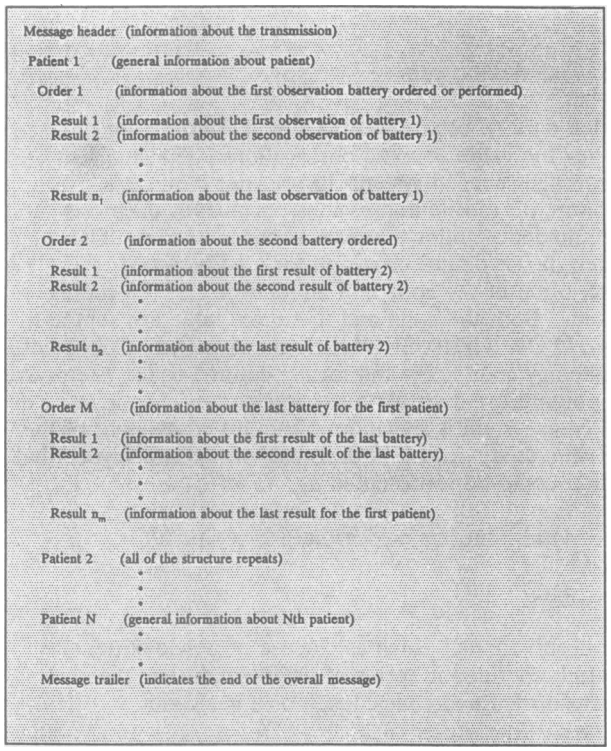
The standard specifies the syntax of the message including content and format. For many data elements, it does not define the semantics (the meanings of codes in individual data elements). It leaves that up to organizations, such as the World Health Organization (ICD9 and ICD10), the American College of Pathology (SNOMED), the European Community (Euclides)[4], and others.

A message consists of a number of different segments that correspond to different record types in a data base. The major segments are 1) the message header, 2) the patient (about whom orders or results are being sent), 3) the specimen/order, and 4) the individual observations or results. The order segment is based on the metaphor of the request form. It carries the order information to a

diagnostic service and then returns this same information plus the results back to the requestor.

An interchange consists of different message segments related in a hierarchy. Figure 1 shows a schematic of a generalized message . Each message contains exactly one message header, which is followed by one or more patient segments. Beneath each patient segment may appear one or more "order" segments, and beneath each "order" segment, one or more observation segment. The results for an SMA12, would be returned as one patient segment, one order (result header) segment and twelve observation segments. If a set of vital signs were also being sent, another order record and separate observation records for Pulse, Respiration rate, Diastolic blood pressure, and Systolic blood pressure, would be required.

Figure 1 Logical Structure of Message



Observations may be sent in both directions between the requesting and the producing service. Thus observations about the patient may be sent to a diagnostic service with an order, for example, the date of the last menstrual period with a cervical pap test order.

The standard does not define separate fields for each observation. There is not one field for sending a glucose result, another for a diastolic blood pressure, and still another for chest tube output. The standard requires a separate "record" for each observation reported. That record has one field (field #4) for identifying the observation and another field (field #6) for reporting its value. In some sense these are "name-value" pairs. When a clinical system wants to transmit information about 6 different observations, it sends 6 observation records, one per value (see example in Figure 2).

Figure 2 Sample ASTM Interpretive Report Messages

The text within these examples that begins with **-- and ends with --** are explanatory comments, not a formal part of the message stream.

--Patient identifying record--
P|1|A9999|P10098|245-83-2033|Hammond^William^Edward^II||19350109|M|W <CR>

--Order record for EKG--
OBR|1|P8753|EK5230|93000^EKG|R|19870329153O|19873290800|||401-0^INTERN&JOE&&&MD^L|N <CR>

--Two interpretation records for EKG--
[In this case, the result observation ID assumes the observation code in the order record.]
OBX|1|CE|&IMP|1|^Sinus bradycardia|||A <CR>
OBX|2|CE|&IMP|2|^Occasional PVCs|||A <CR>

--Four numeric results for EKG--
[The AS4 code is an extension of the CPT4 code (93000) for EKG plus extension .1, .2, etc., as detailed in Appendix A.]
OBX|3|ST|93000.1^Ventricular rate|80|/min|60-100 <CR>
OBX|4|ST|93000.2^Atrial rate|80|/min|60-100 <CR>
OBX|5|ST|93000.4^QRS width|.08|msec|.06-.10 <CR>
OBX|6|ST|93000.3^PR interval|.22|msec|.18-.22 <CR>

Other important information, such as the normal range, the units, and the degree of abnormality of the observation, are included along with the results in the observation (OBX) segment. This approach makes it easy to expand the universe of observations. To add to the universe, you add to the table of observation IDs. There is no need to tinker with the field definitions of the standards.

Message segments are composed of variable length fields. The Patient segment, for example, contains a field for name, hospital number, birth date, and so on (see Table 2 for the complete list for the header, patient, order, and observation segments). The contents of each field is separated from the next by a designated field delimiter. The vertical bar (|) is the recommended field delimiter.

Table 2 Synopsis of Field Names and Their Maximum Lengths

Section	Field Name	Type	Length max
7	Message Header (H)		
7.1	Record type ID	ST	3
7.2	Delimiter definition	ST	5
7.3	Message control ID	ST	12
7.4	Access password	ST	12
7.5	Sender Name or ID	ST	40
7.6	Sender street address, city, state and zip code	AD	100
7.7	Message type	ID	7
7.8	Sender telephone number	TN	40
7.9	Characteristics of sender	ST	40
7.10	Receiver ID	ST	40
7.11	Comment or special instructions	ST	80
7.12	Processing ID	ID	20
7.13	Version	NM	5
7.14	Date and time of message	ST	20
8	Patient Record (P)		
8.1	Record type ID	ST	3
8.2	Transmission sequence number	TN	4
8.3	Practice assigned patient ID	CK	16
8.4	Diagnostic service assigned patient ID	CK	16
8.5	Alterative patient ID	ST	16
8.6	Patient name	PN	48
8.7	Mother's maiden name	ST	24
8.8	Birthdate	TS	12
8.9	Patient sex	ID	1
8.10	Patient race or ethnic origin	ID	40
8.11	Patient street address City, state, country & zip postal code	AD	200
8.12	Not used		
8.13	Patient telephone number	TN	40
8.14	Attending physician ID	CNA	60
8.15	Special field 1	ST	60
8.16	Special field 2	ST	60
8.17	Patient height	CQ	10
8.18	Patient weight	CQ	10
8.19	Patient's known or suspected diagnosis	CE	200
8.20	Patient's medications	ST	200
8.21	Patient's diet	ST	200
8.22	Practice field 1	ST	60
8.23	Practice field 2	ST	60
8.24	Admission date and discharge date	TS	20
8.25	Admission status	ID	2
8.26	Location	ST	25
8.27	Diagnostic classification	ID	10
8.28	Patient religion	ID	30
8.29	Marital status	ID	2
8.30	Isolation status	ID	20
8.31	Language	ST	20
8.32	VIP status	ID	20
8.33	Date Registration Changed	TS	19

Table 2 (continued) Synopsis of Field Names and Their Maximum Lengths

Section	Field Name	Type	Length max
9	Test Order Record (OBR)		
9.1	Record Type ID	ST	3
9.2	Sequence number	NM	4
9.3	Requestor (placer) specimen ID or accession #	CM	75
9.4	Diagnostic service (filler) specimen ID or accession #	CM	75
9.5	Universal battery ID	CE	200
9.6	Priority	ST	2
9.7	Requested date-time	TS	19
9.8	Specimen collection or observation date-time	TS	19
9.9	Collection/observation end time	TS	19
9.10	Collection volume	CQ	20
9.11	Collector ID	CNA	60
9.12	Action code	ST	1
9.13	Danger code	CM	60
9.14	Relevant clinical information	ST	300
9.15	Date and time of specimen receipt	TS	19
9.16	Source of specimen	CM	300
9.17	Ordering provider	CNA	60
9.18	Order call back telephone number	TN	40
9.19	Requestor (placer) field 1	ST	60
9.20	Requestor (placer) field 2	ST	60
9.21	Diagnostic service (filler) field 1	ST	60
9.22	Diagnostic service (filler) field 2	ST	60
9.23	Date and time results reported or status changed	TS	19
9.24	Diagnostic service charge to practice	CM	40
9.25	Diagnostic service section ID	ID	10
9.26	Results (order) status code	ST	1
9.27	Linked results accession #	CM	200
9.28	Quantity - Timing	CM	200
9.29	Result copies to	CNA	150
9.30	Parent access #s	CM	60
9.31	Transportation mode	ID	20
9.32	Reason for study	CE	300
9.33	Assisting interpreter of study (resident)	CNA	60
9.34	Principle interpreter of study records	CNA	60
9.35	Technician identity	CNA	60
9.36	Transcriptionist identity	CNA	60
9.37	Date-time scheduled	TS	19
10	Result Observation Record (OBX)		
10.1	Record type ID	ST	3
10.2	Set ID	NM	4
10.3	Value type	ID	2
10.4	Observation identifier	CE	80
10.5	Observation Sub-ID	ST	20
10.6	Observation value	(Variable)	65K
10.7	Units	ST	20
10.8	Reference range	ST	60
10.9	Abnormal flags	ST	10
10.10	Probability	NM	5
10.11	Nature of abnormal test	ID	5
10.12	Observation result status	ID	2
10.13	Date last obs normal values	TS	19

Some fields may contain repeating values. E.g., the patient name field may contain multiple names corresponding to each of a patients aliases. Repeats must be separated by a repeat delimiter (tilde (~) is the recommended repeat delimiter). A field value may also consist of components. The patient name has six components: last name, first name, middle name and so on. Components are separated by a component designated delimiter, the carat (^) is suggested. Thus, Superman's names would be transmitted as follows:

... |Superman~Kent^Clark|

The standard specifies a number of data types. Fields with the same data type all have the same data representation. Three important ones.

TS = Time Stamp (date -time)

A date is transmitted as YYYYMMDD, according to the ISO date-time standards. A date plus a time is transmitted as YYYYMMDDHHMM. Noon on December 31, 1991, would thus be represented as:

199112311200

There is provision for seconds, fractions of seconds (where needed) and for time zones.

AD = Address

An address consists of the following 6 components, separated by component delimiters:

a) street address or post office box #
b) apartment number or other internal address
c) city
d) state or province
e) zip or postal code
f) country

e.g., 1432 Hosteler Street^Apt 232^Chicago^IL^60603^USA

The country designation is only required when the address is outside of the source country.

CE Coded data type

Coded data types are used to identify observations and to convey the value of some observations. The standard encourages clinical systems to transmit symptoms, findings, diagnostic impressions, problems, diagnoses, and recommended follow-up tests as coded entries. The coded data type has provision for a code, a text description of the code and an identifier of the coding system. Both the code and its source table are sent as data. ICD9, ICD10, SNOMED, the American College of Radiology code book and locally defined code systems are all potential sources for diagnostic codes. The identity of the code system is specified in the 3rd component of the CE data type. An ICD9 diagnosis of anterior myocardial infarct could be represented as:

410.1^anterior myocardial infarction^I9C
or as
410.1^^I9C

An observation ID such as glucose or pelvic outlet size could be defined in terms of CPT4 (with the AS4 extensions), Reed codes, local, or other codes.

By sending the code and its source, we decouple the data interchange standard from the semantics. This allows for the code development to proceed independently of the interchange standard, at its own pace. It eases conversion from one code system (e.g. ICD9) to another (E.g., ICD10) because both kinds of codes can co-exist within messages during the transition stage. It permits the use of different code systems for different purposes, SNOMED, for example, for surgical pathology diagnosis, and ICD9 for discharge diagnoses. It also permits locally coding systems to be used among cooperating sites. A referral laboratory could, for example, identify the test results to its client practices by means of its internal test identifier codes. Local codes are a necessity for the near term because existing universal codes are not comprehensive enough for some applications.

WORKING GROUP	CONTACT
ASTM E31.11: Clinical Data Interchange	Clem McDonald, M.D., Chairman Regenstrief Institute, 1001 W 10th St., Indianapolis, IN 46202; (317) 630 7070
ASTM E31.12: Medical Informatics	Elmer Gabrieli, M.D., Chairman Gabrieli Medical Information, Statler Towers, Suite 1633, Buffalo, NY 14202; (716) 856-2890
ASTM E31.14: Interfaces for Laboratory Instruments	Leon B. Wolf, Chairman E. I. DuPont de Nemours Co., Inc., Medical Products Dept., Concord Plaza, Ridgely Bldg 2A5, Wilmington, DE 19898; (302) 695-5685
ASTM E31.15 Medical Knowledge Representation	Al Pryor, Ph.D., Chairman University of Utah, LDS Hospital, 325 8th Ave, Salt Lake City, UT 84143; (801) 321 2123
ASTM E31.16 Data Exchange for Clinical Neurophysiology	Ernie Jacobs, M.D., Chairman 9500 Euclid Ave., Cleveland, OH 44106; (216) 444 7006
Health Level Seven (HL7): Hospital Network Systems	Ed Hammond, Ph.D. Duke University Medical Center, P O Box 2914, Durham, NC 27710 (919) 684 6421
IEEE MEDIX P1157	Jack Harrington Hewlett Packard, 175 Wyman St., Waltham, MA 02254; (617) 890-6300
ACR/NEMA Radiological Imaging Standards	Bob Thompson, Chairman Dept. of Radiology, Research Div., University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7510; (919) 966-1770
IEEE MEDIX P1073	Jack Harrington Hewlett Packard, 175 Wyman St., Waltham, MA 02254; (617) 890-6300
CEN 001	George DeMoor, Ph.D., Chairman State University Hospital of GENT, Dept. of Medical Informatics, De Pitalaan 185-5K3, B-9000 GENT, Belgium; 32 91 403436

Other Efforts

The early years of standards development were lonely. No longer. Today a variety of groups are developing data interchange standards. For the most part these groups are dealing with non-overlapping domains and/or the development is well coordinated. A very few "border disputes" remain to be resolved, however.

Here I will describe the organizations involved. Health Level 7 (HL7) is a consortium of vendors, users, and consultants, who are developing interchange standards for all of the transactions that occur in a large medical institution. They include admission-discharge-transfer, order entry, result reporting, billing, etc, in their scope. HL7 has been very successful in recruiting information system vendors to the effort, and more than 40 sites are now implementing the HL7 standard. HL7 and ASTM E31.11 are closely allied and coordinated. The two groups have the same philosophy, message structure and data types. The major segments of ASTM 1238 are a subset of the segments in HL7 2.1.

Two sister subcommittees of ASTM E31.11 are also developing data interchange standards. E31.14 is developing standards for linking laboratory instruments to laboratory systems. They have completed the balloting on two standards, one for the lower level protocol and the other for the content of the interchange message. The latter follows the E31.11 formats closely. Both standards are likely to be published by the time of SCAMC '90. E31.16 is developing standards for transmitting signals and wave forms from physiologic monitors. There first task group has focussed on neurophysiologic signals. This standard uses the segments and the general framework of ASTM 1288, but adds new conventions for sending the numeric value of points in the wave form.

The American College of Radiology in conjunction with the national association of manufacturers (ACR/NEMA) published a standard for the interchange of radiologic images in 1985. This has been adopted by most PACS vendors, and is likely to become an international standard. ACR/NEMA is working currently on standards for order entry and result reporting.

IEEE has two data interchange efforts. The Medical Information Bus (MIB) committee has developed (and is about to publish) interchange standards for critical care devices such as physiologic monitors, automatic intravenous infusion pumps and the like. The standard includes specifications at the hardware and the logical level. MEDIX, another IEEE committee, has a very broad charter. They intend to bring all nodalities of all clinical data interchanges within the ISO scope. They will use HL7 as the initial message definition.

Because standards will be vital to a successful economic union, the European community is vigorously developing standards for a wide variety of issues, among which are standards for clinical data interchange. The European effort is being coordinated by the CEN 001 committee under the leadership of Dr. George DeMoor of State University Hospital, Gent (Belgium). Euclides, one of their completed projects focuses on the transmission of clinical laboratory data. The project team has developed a message structure and proposals for the lower level protocol as well as a bridge to ASTM 1238. They have also developed a comprehensive coding system for test identifiers, a major, and much needed, contribution.

Conclusion

Our goal is to standardize the communication of clinical data between clinical systems, not the systems themselves nor their internal operation. In fact, doing the latter would be a mistake at this point in history. It would deflect energy from, and delay the spread of CDI standards. But worse, it keeps attention on the systems where it has been, instead of on the data where it should be. The data is the most expensive part of any data system. It is also the reason de etre for such systems. Yet we accept our clinical data being like the boy in the bubble. Our data generally cannot "live" outside of the computer system in which it was born. Electronic data should not depend upon the internals of a particular program, or language, or machine, for its interpretation. We need to be able to look at the recorded data from 2 years ago and be able to compare it to data from today. (Whoops, the vendor has changed the file structure so we cannot look at it, at all).

The data interchange standards give our data life independent of the source system. There are two components to such standards, the message format or syntax, and the dictionary of codes (semantics). Much of this is now available, or on the way. For many applications, message standards are here. For a few kinds of clinical entities, e.g., drugs, the code systems (e.g., the National Drug Code) are virtually complete. The

available universal codes for clinical descriptors are not adequate. The National Library of Medicine UML project will reduce this deficiency, but there is much to be done. In this country, our codes for clinical variables, such as blood pressure, and blood glucose that vary regarding method, units, normal ranges, and physiologic correlates, are inadequate. Work being done by ASTM E31.12 and the Euclides project offers promise for clinical laboratory values. If we want to pool data from different institutions for clinical and policy research, however, this area needs urgent attention.

Ancillary service systems that capture data are like Johnny one-notes. They can play a few notes on one instrument. To care for a patient we need to hear all of the notes of all of the instruments. The MIDI standard lets one musician combine many notes from one instrument into songs and then combine the songs. So, one person can then create a concerto or a symphony. Care givers need the same power.

References:

1) Aikin J.: Musical instrument digital interface. Keyboard Magazine. 1986;1:28-31.

2) McDonald CJ: The search for national standards for medical data exchange. M.D. Computing. 1984;(1)1:3-4.

3) A Standard Specification for Transferring Clinical Laboratory Data Messages Between Independent Computer Systems. ASTM 1238-88. Philadelphia: American Society for Testing Materials; 1988.

4) Euclides, a European standard for clinical laboratory data exchange. Readings in Medical Informatics. ISBN 90-73045-03-7, pg 373-375. Editor, G. DeMoor.

Supported by grants (HS-04996 and HS5626) from the Agency for Health Care Policy and Research